ADROID Conference Series: Engineering and Technology
Volume 2, Issue 3, pp.169-181, 2026

NeuroGuard: Digital Eye Fatigue Monitoring via Hybrid Landmark Analysis

Agnideep Poddar, Kshitij Gupta
Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Chennai, India
ap2774@srmist.edu.in, kg3007@srmist.edu.in

Abstract. Fatigue has a pronounced effect on reaction times, sustained attention, decision-making, and visual performance, and is a serious threat to road traffic safety and occupational productivity. Therefore, real-time and reliable eye fatigue detection systems are key components of intelligent transportation and human-computer interaction technologies. Existing vision-based methods either rely solely on deep learning to extract spatial features without interpretable physiological markers or require manually designed geometric features with conventional classifiers, which have limitations.
In this paper, we propose an efficient real-time eye fatigue monitoring system, called NeuroGuard, based on a hybrid Convolutional Neural Network–Long Short-Term Memory (CNN–LSTM) deep learning architecture and MediaPipe-based facial landmark tracking. The system combines spatial feature representations with temporal eye behavior patterns to provide a method for quantifying blink rate and eyelid closure duration using Eye Aspect Ratio (EAR) and Percentage of Eye Closure (PERCLOS) metrics. Experimental evaluation in three real-world scenarios, including professional driving, long-term software development, and night-shift monitoring under low-light conditions, demonstrates that NeuroGuard achieves 94.3% accuracy, 96.2% recall, and 91.8% precision, improving by 12.2%, 11.2%, and 13.4%, respectively, over traditional threshold-based baselines. Our system reduces false positives from 14.5% to 3.2% and achieves a practical inference latency of approximately 13 ms, making it suitable for safety-critical real-time deployment.

Keywords: Artificial Intelligence, CNN–LSTM, MediaPipe, Eye Aspect Ratio (EAR), Percentage of Eye Closure (PERCLOS), Real-Time Monitoring, Driver Drowsiness Detection, Digital Eye Fatigue

1. INTRODUCTION

Eye fatigue and drowsiness have become major issues in modern health and safety, driven by increased screen time, long-distance driving, and extended virtual reality interaction. Fatigue-related factors are involved in approximately 20% of serious road crashes worldwide, according to the World Health Organization. Fatigue significantly reduces cognitive capacity, slows reaction times, decreases concentration, and impairs decision making, all of which are essential for safe and productive human performance [1,2].
Visual signs are among the most reliable physiological indicators of fatigue. Research has established strong correlations between fatigue levels and blink frequency, eyelid closure duration, gaze deviation, pupil dynamics, and eyelid movement velocity [3,4]. These markers enable non-contact fatigue detection using standard cameras, making vision-based fatigue monitoring practical and scalable across transportation, industrial environments, and workplace applications.
Advances in computer vision and artificial intelligence have accelerated the development of increasingly sophisticated non-contact fatigue detection systems. Early studies employed classical machine learning classifiers with geometric ocular features, including lightweight Support Vector Machine (SVM)-based digital eye strain detection [1], multi-feature fusion models incorporating facial behavioral characteristics [2], temporal sequence modeling techniques [3], and blink-pattern-based fatigue prediction [4].
With the emergence of deep learning, Convolutional Neural Networks (CNNs) substantially improved eye-state classification by automatically learning spatial visual features [5]. Transfer learning approaches enhanced robustness across different facial poses [6], while multi-frame analysis and Gaussian-weighted techniques improved performance under real-world conditions [7,8]. Additionally, gaze-based models have been explored for fatigue detection during interactive tasks [10], and electroencephalography (EEG)-based neurosignal analysis has been investigated for estimating cognitive load and eye fatigue [11].
Despite these advances, existing research reveals a persistent limitation. Geometry-based approaches are computationally efficient and interpretable but often fail to capture the temporal progression of fatigue. Conversely, deep learning models achieve higher predictive accuracy but generally lack physiological interpretability and may introduce higher computational complexity [12]. Furthermore, many existing studies

ISBN: 978-81-995593-3-2

evaluate only a single modeling approach, rarely reporting inference latency or interpretability alongside performance metrics.

The primary motivations for this work are:

1. The inability of single-feature methods to adequately represent the multidimensional characteristics of fatigue.
2. The lack of hybrid systems that integrate physiologically interpretable features with deep temporal learning.
3. The need for robust fatigue detection systems capable of operating reliably under varying illumination conditions and natural head movements.

The major contributions of this work are summarized below:

- Development of a hybrid CNN–LSTM architecture that jointly models spatial eye-region features and temporal blink dynamics for accurate drowsiness classification.
- Integration of physiologically interpretable fatigue biomarkers, namely Eye Aspect Ratio (EAR) and Percentage of Eye Closure (PERCLOS), extracted from MediaPipe Face Mesh using 468 facial landmarks.
- Experimental validation across three practical scenarios:
    - Professional long-haul driving
    - Extended software development (cognitive fatigue)
    - Low-light night-shift monitoring
- Design of a modular real-time monitoring framework incorporating video acquisition, fatigue analysis, database storage, and interactive dashboard visualization.

The remainder of this paper is organized as follows. Section 2 presents the literature review. Section 3 describes the proposed system architecture. Section 4 explains the methodology. Section 5 presents the performance metrics used for evaluation. Section 6 discusses the experimental evaluation and results. Section 7 outlines the limitations and future work. Finally, Section 8 concludes the paper.

2. LITERATURE REVIEW

Vision-based eye fatigue detection has evolved significantly, progressing from traditional geometric feature extraction techniques to advanced hybrid deep learning architectures. Existing research can be broadly categorized into geometric and statistical methods, deep learning-based spatial approaches, hybrid spatial-temporal models, and emerging multimodal techniques. A comparative summary of representative studies is presented in Table 1.

2.1 Geometric and Statistical Feature-Based Methods

Early eye fatigue detection systems primarily relied on handcrafted geometric features and conventional machine learning algorithms. Guleria and Kaur proposed a lightweight digital eye strain detection system based on Support Vector Machines (SVMs) using handcrafted geometric eye features, achieving an accuracy of approximately 88%. Although computationally efficient and highly interpretable, the method was unable to capture the temporal progression of fatigue.

Qin *et al.* introduced a multi-feature empirical fusion model that combined blink rate, mouth movement, and head position, improving classification accuracy to approximately 90%. Similarly, Xiang employed Gabor wavelet-based feature extraction together with Hidden Markov Models (HMMs) for temporal fatigue recognition. Kuwahara *et al.* further demonstrated that changes in blink frequency serve as reliable physiological indicators of accumulated fatigue, reinforcing the effectiveness of blink-based monitoring techniques.

2.2 Deep Learning-Based Spatial Methods

The emergence of deep learning significantly enhanced vision-based fatigue detection by enabling automatic extraction of discriminative visual features. Ni demonstrated that Convolutional Neural Networks (CNNs) effectively classify open and closed eye states with an accuracy of approximately 89%.

Madni *et al.* applied transfer learning using pre-trained ResNet models to improve robustness against pose variations, achieving approximately 91% accuracy. Xie and Fan proposed a multi-frame visual feature framework for real-time fatigue detection, while Xiang *et al.* introduced a Gaussian-weighted pixel distribution technique that improved the stability of eye-state recognition. Kayadibi *et al.* further demonstrated the effectiveness of

AlexNet-based deep CNNs, highlighting the advantages of transfer learning and pre-trained architectures for eye-state classification.

2.3 Hybrid Spatial–Temporal Approaches

Hybrid CNN–LSTM architectures have effectively addressed the limitations of purely spatial models by incorporating temporal information into fatigue prediction. Selvi *et al.* developed a CNN–LSTM model for digital screen users, achieving an accuracy of approximately 92%, although its evaluation was limited to controlled environments.

Lv and Pei proposed a Bidirectional Long Short-Term Memory (BiLSTM) framework combined with a comprehensive fatigue index for driver fatigue prediction. Lian *et al.* introduced a hybrid EEG and eye-tracking system that significantly improved detection accuracy but required specialized hardware, reducing its practicality for widespread deployment. Wang *et al.* proposed a temporal facial feature-based fatigue detection system optimized for low-cost devices, demonstrating the growing interest in efficient real-time implementations.

2.4 Emerging and Multimodal Approaches

Recent research has increasingly focused on multimodal fatigue detection by combining multiple physiological signals. Sidaq *et al.* proposed the Virtual Meeting Fatigue Detector (VMFD), which estimates fatigue using polygon-area analysis of eye activity during virtual meetings.

Chae *et al.* designed an EEG–SSVEP-based interface that reduces visual fatigue through selective attention mechanisms. Blanco-Díaz *et al.* investigated the influence of eye fatigue on Brain–Computer Interface (BCI) performance based on ERP-P300 signals. Sun *et al.* explored various eye-tracking signal features for detecting screen-use fatigue, demonstrating the growing importance of multimodal sensing in fatigue assessment.

2.5 Research Gaps

Despite considerable advancements, several important challenges remain in the existing literature. First, most studies evaluate only a single modeling approach without conducting comprehensive comparisons against alternative methods. Second, practical deployment aspects such as model interpretability and inference latency are rarely reported alongside detection accuracy. Third, relatively few studies validate their methods under challenging real-world conditions, including low-light environments and natural head movements. Finally, integrated systems that combine physiologically interpretable biomarkers with deep temporal learning remain limited.

The proposed NeuroGuard framework addresses these limitations by integrating interpretable physiological indicators, namely Eye Aspect Ratio (EAR) and Percentage of Eye Closure (PERCLOS), with a hybrid CNN–LSTM architecture capable of learning both spatial eye features and temporal fatigue dynamics. Furthermore, the system is validated across multiple real-world scenarios, demonstrating its practicality for robust real-time fatigue monitoring.

TABLE 1. COMPARATIVE LITERATURE SUMMARY

Reference	Method	Key Feature	Accuracy	Limitation
Guleria and Kaur	SVM with Geometric Features	Eye HOG + SVM	~88%	No temporal modeling
Qin *et al.*	Multi-feature Fusion	Blink, mouth movement, head position	~90%	High computational cost
Ni	CNN Eye-State Classification	Spatial CNN features	~89%	No temporal sequence modeling
Madni *et al.*	Transfer Learning	Pre-trained CNN (ResNet)	~91%	Sensitive to pose variations
Proposed NeuroGuard	Hybrid CNN–LSTM	EAR, PERCLOS, Temporal Features	~94.3%	Reduced performance for head rotations greater than 45°

3. SYSTEM ARCHITECTURE

The NeuroGuard framework is designed as a modular and loosely coupled architecture in which each component performs a dedicated function, including video acquisition, physiological feature extraction, deep learning-based fatigue classification, persistent data storage, and interactive visualization. This modular design improves scalability, maintainability, and flexibility, enabling individual modules to be upgraded or replaced without affecting the overall system. The complete architecture of the proposed framework is illustrated in Figure 1.

3.1 Real-Time Acquisition Layer

The Real-Time Acquisition Layer is responsible for continuously capturing RGB video streams using a standard webcam interface. The captured frames are preprocessed through the MediaPipe Camera Utilities before further analysis. During preprocessing, each frame is resized to a fixed resolution, luminance is normalized, and contrast is adjusted to ensure consistent image quality. These preprocessing operations reduce the influence of environmental lighting variations and differences in camera hardware, thereby improving the robustness of facial landmark detection. In addition, adaptive histogram equalization is applied to minimize shadow and glare artifacts under varying illumination conditions, enabling reliable fatigue monitoring in diverse real-world environments.

3.2 Processing Engine

The Processing Engine forms the core analytical component of the NeuroGuard system and operates in two sequential stages.

In the first stage, MediaPipe Face Mesh extracts 468 three-dimensional facial landmarks from every preprocessed video frame. These landmarks include highly accurate periocular keypoints that precisely define the eyelid boundaries. Based on these landmark coordinates, two physiologically interpretable fatigue indicators are calculated:

- Eye Aspect Ratio (EAR), which measures the degree of eyelid openness.
- Percentage of Eye Closure (PERCLOS), which quantifies the proportion of time during which the eyes remain substantially closed within a sliding temporal window.

The computation of these metrics is described in detail in the Methodology section.

In the second stage, the temporal sequence of EAR and PERCLOS values is combined with cropped eye-region images and provided as input to the proposed hybrid CNN–LSTM network. The CNN component extracts discriminative spatial features from eye images, while the LSTM component models temporal dependencies across consecutive frames to capture blink patterns and fatigue progression. Based on these learned spatial-temporal representations, the system performs real-time classification of the user's condition into one of three states:

- Alert
- Drowsy
- Fatigued

3.3 Persistence Layer

The Persistence Layer stores fatigue analysis results in a PostgreSQL relational database using the Drizzle Object Relational Mapping (ORM) framework. Each record includes timestamps together with the computed fatigue indicators and classification outcomes. This persistent storage supports both immediate alert generation and long-term analysis of user fatigue trends. Historical records enable retrospective assessment of fatigue patterns and facilitate personalized threshold calibration based on an individual's baseline Eye Aspect Ratio (EAR) and blink behavior, thereby improving long-term monitoring accuracy.

3.4 Presentation Layer

The Presentation Layer provides an interactive Human–Machine Interface (HMI) developed using React and Vite, while communication between the client and server is handled through an Express.js REST API. Real-time visualization of fatigue metrics is implemented using the Recharts visualization library, which displays continuous time-series graphs of Eye Aspect Ratio (EAR), PERCLOS values, and fatigue classification outputs.

The dashboard provides immediate visual and auditory alerts whenever drowsiness or fatigue is detected. It also allows users and supervisors to configure alert thresholds and examine historical fatigue trends, thereby supporting both real-time monitoring and long-term performance analysis.

Figure 1. System Architecture of NeuroGuard

4. METHODOLOGY

The proposed NeuroGuard framework follows a structured three-stage detection pipeline comprising (1) Data Acquisition and Preprocessing, (2) Physiological Feature Engineering, and (3) Hybrid CNN–LSTM Modeling. This pipeline combines interpretable physiological measurements with deep learning-based temporal analysis to accurately identify eye fatigue in real time.

4.1 Data Acquisition and Preprocessing

High-resolution RGB video frames are continuously captured using a standard webcam. To ensure temporal consistency throughout the monitoring process, the video stream is maintained at a stable frame rate. Every captured frame undergoes a series of preprocessing operations before feature extraction.

Initially, each frame is resized to a fixed resolution of 224 × 224 pixels to provide a uniform input size for the deep learning model. Channel-wise normalization, including mean subtraction and standard deviation scaling, is then performed to standardize pixel intensity values. Finally, adaptive histogram equalization is applied to reduce the effects of shadows, glare, and varying illumination conditions, thereby improving image quality under different lighting environments.

Facial landmark detection is subsequently performed using the MediaPipe Face Mesh framework, which extracts 468 three-dimensional facial landmarks from every frame, generating a dense representation of the facial surface. To reduce sensitivity to variations in face-to-camera distance and minor head movements, all landmark coordinates are L2-normalized into unit vectors. This normalization step enhances cross-subject consistency and improves the generalization capability of the fatigue detection system.

4.2 Feature Engineering

The proposed framework combines interpretable physiological biomarkers with deep learning representations to improve fatigue detection accuracy. Two clinically validated eye fatigue indicators are extracted from the facial landmarks: Eye Aspect Ratio (EAR) and Percentage of Eye Closure (PERCLOS).

4.2.1 Eye Aspect Ratio (EAR)

The Eye Aspect Ratio (EAR) measures the geometric openness of the eye by calculating the ratio between the vertical and horizontal Euclidean distances among six periocular landmarks. It is computed as:

$$\mathrm{EAR} = \frac{\| p_2 - p_6 \| + \| p_3 - p_5 \|}{2 \| p_1 - p_4 \|}$$

where:

- p_1 and p_4 denote the horizontal corner landmarks of the eye (medial and lateral canthus).
- p_2, p_3, p_5, and p_6 represent the upper and lower eyelid landmarks.

Under normal conditions, the EAR typically ranges between 0.30 and 0.40 when the eyes are fully open and approaches zero when the eyes are closed. A sustained reduction in EAR below a calibrated threshold of approximately 0.20 indicates prolonged eyelid closure and is considered an indicator of drowsiness.

4.2.2 Percentage of Eye Closure (PERCLOS)

The Percentage of Eye Closure (PERCLOS) measures the proportion of time during which the eyes remain at least 80% closed within a 60-second sliding observation window. It is calculated as

$$\text{PERCLOS} = \frac{\text{Number of frames with } EAR \leq 0.2 \times EAR_{open}}{\text{Total number of frames}}$$

PERCLOS is a clinically validated fatigue biomarker recommended by the Federal Highway Administration (FHWA) for transportation safety research. A PERCLOS value greater than 0.15, indicating that the eyes remain closed for more than 15% of the observation period, is generally regarded as a strong indicator of significant fatigue.

Both EAR and PERCLOS are computed for every video frame and organized into temporal sequences, which serve as inputs to the hybrid deep learning model.

4.3 Hybrid CNN–LSTM Architecture

The proposed hybrid CNN–LSTM architecture simultaneously captures spatial characteristics of the eye region and temporal patterns associated with fatigue progression.

The CNN component receives cropped eye-region images of size 64 × 64 pixels as input. It consists of three consecutive convolutional blocks, each comprising:

- A 3 × 3 convolutional layer
- ReLU activation
- Batch Normalization
- Max-Pooling

These convolutional layers extract discriminative visual features related to eye openness and eyelid movement. The resulting feature maps are flattened to generate a 512-dimensional feature vector for each frame.

The extracted visual features are then augmented with the corresponding EAR and PERCLOS values. Feature vectors from 30 consecutive frames (approximately one second of video at 30 frames per second) are stacked to form a temporal sequence.

This sequence is processed by a two-layer Long Short-Term Memory (LSTM) network containing 256 hidden units in each layer. A dropout rate of 0.4 is applied to reduce overfitting. The LSTM captures long-term temporal dependencies in blink behavior, enabling the model to recognize gradual reductions in blink frequency and increasing eyelid closure duration that typically precede fatigue.

The output from the final LSTM layer is passed through a fully connected classification layer followed by a Softmax activation function, producing binary predictions corresponding to either Alert or Drowsy.

Model training is performed using the Binary Cross-Entropy Loss function and the Adam optimizer with a learning rate of 3×10^{-4}. The network is trained for 60 epochs, while early stopping with a patience value of 10 epochs is employed to prevent overfitting and improve generalization performance.

5. PERFORMANCE METRICS

The performance of the proposed NeuroGuard classifier is evaluated using five standard classification metrics: Accuracy, Precision, Recall (Sensitivity), F1-Score, and False Positive Rate (FPR). In this study, drowsiness is considered the positive class. The evaluation metrics are computed using the following confusion matrix components:

- TP (True Positives): Correctly classified drowsy instances.
- TN (True Negatives): Correctly classified alert instances.
- FP (False Positives): Alert instances incorrectly classified as drowsy.
- FN (False Negatives): Drowsy instances incorrectly classified as alert.

These metrics collectively assess the effectiveness, reliability, and practical usability of the proposed fatigue detection system.

5.1 Accuracy

Accuracy represents the overall proportion of correctly classified samples among all observations. It provides a general measure of the classifier's performance by considering both correctly detected drowsy and alert states.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

A higher accuracy indicates that the model correctly identifies both fatigued and non-fatigued conditions across the dataset.

5.2 Precision

Precision, also known as Positive Predictive Value (PPV), measures the proportion of predicted drowsy instances that are drowsy. High precision is desirable because it minimizes unnecessary fatigue warnings and reduces the occurrence of false alarms that may inconvenience users.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Higher precision indicates greater confidence in the fatigue alerts generated by the system.

5.3 Recall (Sensitivity)

Recall, or Sensitivity, measures the ability of the classifier to correctly identify actual drowsy states. It represents the percentage of true fatigue events that are successfully detected by the system.

$$\text{Recall} = \frac{TP}{TP + FN}$$

A high recall value is particularly important in fatigue monitoring because missed detections (false negatives) can result in serious safety risks, especially in transportation and industrial applications.

5.4 F1-Score

The F1-Score is the harmonic mean of Precision and Recall. It provides a balanced evaluation of the classifier, particularly when the dataset is imbalanced or when both false positives and false negatives carry significant consequences.

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

A higher F1-Score indicates that the classifier simultaneously achieves high precision and high recall, making it a reliable indicator of overall classification performance.

5.5 False Positive Rate (FPR)

The False Positive Rate (FPR) measures the proportion of alert instances that are incorrectly classified as drowsy. This metric reflects the frequency of false alarms generated by the system.

$$\text{FPR} = \frac{FP}{FP + TN}$$

A lower False Positive Rate improves user confidence in the fatigue monitoring system by reducing unnecessary alerts and preventing alarm fatigue. Maintaining a low FPR is particularly important for real-time safety-critical applications where excessive false warnings can reduce user trust and system acceptance.

6. EXPERIMENTAL EVALUATION AND RESULTS

6.1 Experimental Setup

To comprehensively evaluate the effectiveness of the proposed NeuroGuard framework, three real-world experimental scenarios were designed, representing different operational environments, fatigue types, and monitoring durations ranging from 120 to 240 minutes. Fatigue-related physiological metrics were recorded at a sampling rate of 1 Hz throughout each experiment.

Ground-truth fatigue labels were established through manual annotation by a trained human observer based on observable behavioral indicators, including head nodding, self-reported fatigue levels, and microsleep events. For a rigorous performance comparison, the proposed CNN–LSTM hybrid model and a conventional EAR threshold-based method (EAR threshold = 0.20, eye closure duration > 2 consecutive frames) were evaluated under identical experimental conditions.

6.2 Case Study 1: Long-Distance Logistics (Professional Driver)

The first experiment simulated a 120-minute long-distance driving session using a heavy-duty vehicle simulator operated by a professional driver.

During the initial 60 minutes, the Eye Aspect Ratio (EAR) remained stable at approximately 0.32, indicating that the driver maintained a fully alert condition. However, after approximately 75 minutes, the EAR gradually decreased, indicating progressive fatigue accumulation. Three independent microsleep episodes were observed during the experiment, each characterized by an EAR below 0.15 sustained for more than 2.5 seconds, satisfying established physiological criteria for microsleeping.

The proposed CNN–LSTM model generated a "Drowsy" warning at approximately 80 minutes, nearly seven minutes before the first microsleep episode occurred. As fatigue intensified, the system subsequently upgraded the classification to "Critical Fatigue." This early detection capability resulted from the LSTM's ability to learn gradual temporal changes in the EAR signal rather than relying solely on instantaneous threshold values.

In contrast, the conventional EAR threshold-based approach generated an alert only after the microsleep event had already occurred because it depended exclusively on static threshold violations without considering temporal trends.

Figure 2 illustrates the temporal progression of fatigue accumulation throughout the driving session.

Figure 3 presents the corresponding EAR signal, highlighting progressive drowsiness and microsleep events.

Figure 2. Time-Series Representation of Fatigue Accumulation with Event Annotations (Scenario A: Logistics).

Figure 3. Temporal EAR Signal Depicting Drowsiness and Micro-Sleep Indicators.

6.3 Case Study 2: Extended Software Development (Cognitive Fatigue)

The second experiment investigated cognitive fatigue during an 180-minute software development session. Unlike sleep deprivation, cognitive fatigue primarily affects visual attention and concentration while not necessarily producing prolonged eye closure.

During the initial phase of the experiment, the subject maintained a blink rate of approximately 22 blinks per minute, reflecting voluntary blink suppression associated with sustained concentration. Beyond approximately 110 minutes, the blink rate declined significantly to approximately 6–8 blinks per minute, indicating the onset of staring fatigue, a condition characterized by prolonged visual fixation without substantial eyelid closure.

The conventional EAR threshold method failed to detect this form of fatigue because the subject's eyelid closure duration remained within normal limits throughout the experiment. In contrast, the CNN–LSTM model successfully recognized the gradual reduction in blink frequency and generated fatigue warnings at approximately 115 minutes.

These findings demonstrate the importance of temporal sequence modeling for distinguishing cognitive fatigue from conventional drowsiness, which cannot be achieved through static geometric thresholds alone.

Figure 4 illustrates the temporal variation in blink rate observed during the software development session.

Figure 4. Time-Series Analysis of Blink Rate (bpm) Highlighting Adaptation and Fatigue Effects (Scenario B: Engineering).

6.4 Case Study 3: Night-Shift Monitoring (Low-Light Conditions)

The third experiment evaluated NeuroGuard under challenging low-light conditions representative of industrial night-shift environments. The monitoring session lasted 240 minutes, with ambient illumination maintained below 20 lux.

Despite the adverse lighting conditions, the MediaPipe-based facial landmark detection pipeline maintained an average detection accuracy of approximately 94%. The system effectively suppressed false eye-closure detections caused by head rotations and voluntary eye-rubbing, artifacts that frequently produce false alarms in conventional intensity-based eye detection techniques.

The robustness observed under low illumination is primarily attributed to MediaPipe's model-based facial landmark estimation, which relies on learned facial geometry rather than raw pixel intensity, making it significantly more resilient to image noise and poor lighting conditions.

6.5 Comparative Performance Analysis

The overall quantitative comparison between the traditional EAR threshold-based approach and the proposed NeuroGuard CNN–LSTM model is presented in Table 2.

TABLE 2. COMPARATIVE PERFORMANCE OF TRADITIONAL EAR THRESHOLDING AND THE PROPOSED HYBRID CNN–LSTM MODEL

Metric	Traditional EAR	Proposed Hybrid Model	Improvement
Accuracy	82.1%	94.3%	+12.2%
Precision	78.4%	91.8%	+13.4%
Recall	85.0%	96.2%	+11.2%
F1-Score	81.6%	93.9%	+12.3%
False Positive Rate	14.5%	3.2%	−11.3%

The proposed hybrid CNN–LSTM framework consistently outperformed the conventional EAR threshold-based method across all evaluation metrics.

The most significant improvement was observed in Precision, which increased by 13.4%. This improvement indicates that the LSTM successfully distinguishes genuine fatigue events from natural prolonged blinks, transient gaze deviations, and voluntary eye closures, thereby substantially reducing false fatigue alarms.

Another notable improvement was the reduction of the False Positive Rate (FPR) from 14.5% to 3.2%. Such a reduction is particularly important for practical fatigue monitoring applications because excessive false alarms reduce user confidence and may eventually lead to alarm fatigue.

Figure 5 compares several hybrid deep learning architectures evaluated during the experimental study.

Figure 5. *Comparison of Hybrid Model Architectures for Eye Fatigue Detection.*

6.6 Discussion

The experimental evaluation demonstrates the effectiveness of integrating temporal deep learning with physiologically interpretable fatigue biomarkers. Unlike conventional threshold-based methods, the proposed CNN–LSTM architecture successfully differentiates fatigue-induced eye closure from voluntary blinking by learning temporal eye-behavior patterns over consecutive video frames.

The substantial reduction in the False Positive Rate from 14.5% to 3.2% is primarily attributed to the LSTM network's ability to capture contextual blink dynamics rather than relying solely on instantaneous Eye Aspect Ratio measurements.

Furthermore, NeuroGuard maintained consistently high performance across all three experimental scenarios, including professional driving, prolonged cognitive work, and low-light industrial monitoring, demonstrating strong robustness under diverse operating conditions.

Although the hybrid deep learning model introduces an additional inference latency of approximately 13 ms, the resulting improvements in classification accuracy, early fatigue detection, and false alarm reduction significantly outweigh this computational overhead. Consequently, the proposed framework is well suited for deployment in safety-critical applications such as intelligent transportation systems, industrial worker monitoring, and prolonged screen-based activities. Although the hybrid deep learning model introduces an additional inference latency of approximately 13 ms, the resulting improvements in classification accuracy, early fatigue detection, and false alarm reduction significantly outweigh this computational overhead. Consequently, the proposed framework is well suited for deployment in safety-critical applications such as intelligent transportation systems, industrial worker monitoring, and prolonged screen-based activities.

7. LIMITATIONS AND FUTURE WORK

Although the proposed NeuroGuard framework demonstrates robust performance across a wide range of operational scenarios, several limitations should be acknowledged before large-scale deployment.

7.1 Head Rotation Sensitivity

The accuracy of facial landmark tracking decreases when the user's head rotates beyond approximately 45° relative to the camera. Under extreme head poses, the MediaPipe Face Mesh model is unable to consistently estimate periocular landmark positions, introducing noise into the Eye Aspect Ratio (EAR) measurements and potentially reducing detection accuracy. Future research should investigate three-dimensional head-pose estimation, pose correction algorithms, and multi-camera configurations to improve robustness under large head rotations.

7.2 Occlusion Sensitivity

The proposed system experiences reduced performance when the eye region is partially or completely occluded by objects such as prescription glasses, sunglasses, hair, or other accessories. These occlusions interfere with accurate landmark localization and consequently affect fatigue estimation. Future work should focus on developing occlusion-resistant facial landmark detection methods and image inpainting techniques to improve reliability under such conditions.

7.3 Subject-Specific Calibration

The Eye Aspect Ratio (EAR) threshold employed for fatigue detection was empirically determined using population-level observations. However, natural variations in eye anatomy, age, facial structure, and the use of corrective lenses can produce significant differences in resting EAR values among individuals. Consequently, personalized calibration procedures may be required to maximize detection sensitivity and specificity for individual users. Future studies should investigate adaptive threshold estimation and personalized learning techniques to accommodate inter-subject variability.

7.4 Inference Latency on Edge Hardware

Compared with conventional EAR threshold-based methods, the proposed hybrid CNN–LSTM model introduces an additional inference latency of approximately 13 ms. Although acceptable for most real-time applications, deployment on resource-constrained embedded platforms, such as in-vehicle Electronic Control Units (ECUs) and edge devices, may require further optimization. Techniques including model compression, network pruning, quantization, and knowledge distillation should be explored to reduce computational overhead while maintaining detection accuracy.

7.5 Dataset Scope

The experimental evaluation presented in this work was conducted using three custom real-world case studies rather than widely recognized public benchmark datasets such as NTHU Driver Drowsiness Detection (NTHU-DDD) or the Closed Eyes in the Wild (CEW) dataset. Consequently, direct quantitative comparison with existing state-of-the-art methods remains limited. Future research should validate the proposed framework on publicly available benchmark datasets to facilitate comprehensive performance comparisons and improve reproducibility.

Future Work

Several research directions can further enhance the capabilities of the NeuroGuard framework:

- Integration of gaze-tracking features to complement EAR-based fatigue estimation and improve early fatigue detection.
- Incorporation of multimodal physiological signals, such as Electroencephalography (EEG), Electrooculography (EOG), or heart-rate measurements, to improve robustness in clinical and safety-critical environments.
- Development of unsupervised, self-supervised, or few-shot learning techniques for rapid subject-specific calibration without requiring extensive labeled data.
- Optimization of the proposed framework for edge computing platforms through lightweight neural network architectures and efficient model deployment techniques, enabling real-time fatigue monitoring on low-power embedded devices.

8. CONCLUSION

This paper presented NeuroGuard, a modular real-time digital eye fatigue monitoring system that integrates MediaPipe facial landmark tracking with a hybrid CNN–LSTM deep learning architecture. The proposed framework combines spatial eye-region feature extraction, temporal modeling of blink dynamics, and clinically interpretable physiological fatigue indicators, namely Eye Aspect Ratio (EAR) and Percentage of Eye Closure (PERCLOS), into a unified fatigue detection pipeline suitable for real-time safety-critical applications.

The effectiveness of the proposed system was validated through three representative real-world case studies, including professional long-distance driving, extended software development involving cognitive fatigue, and night-shift monitoring under low-light conditions. Experimental results demonstrated the robustness of NeuroGuard across diverse operational environments, achieving an overall classification accuracy of 94.3%, recall of 96.2%, and precision of 91.8%. Compared with the conventional EAR threshold-based approach, the proposed framework reduced the False Positive Rate (FPR) from 14.5% to 3.2% while maintaining a real-time inference latency of approximately 13 ms.

The experimental findings further highlight the importance of temporal sequence modeling for reliable fatigue detection. The LSTM component effectively captures gradual changes in blink behavior and eyelid dynamics, enabling early recognition of both conventional drowsiness and cognitive fatigue that cannot be identified using static geometric thresholding alone. Furthermore, incorporating temporal context substantially reduces false alarms by distinguishing voluntary prolonged blinks from fatigue-induced eye closures associated with imminent microsleep events.

The modular design of NeuroGuard also provides considerable flexibility for future enhancements, including integration with additional fatigue indicators, personalized user calibration, and deployment on resource-constrained embedded and edge computing platforms. Future research will focus on improving robustness to extreme head rotations and occlusions, developing adaptive subject-specific calibration methods, and validating the proposed framework using publicly available benchmark datasets to facilitate comprehensive comparison with existing state-of-the-art fatigue detection approaches.

Conflict of Interest

The authors declare no conflicts of interest regarding the current research.

References

[1] A. Guleria and R. Kaur, "Digital Eye Strain Detection System Based on SVM," in *Proceedings of the 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, June 2021, pp. 1–6. doi: 10.1109/ICOEI51242.2021.9453085.

[2] Y. Qin, H. Lyu, and K. Zhu, "Driver fatigue detection method based on multi-feature empirical fusion model," *Multimedia Tools and Applications*, vol. 83, no. 14, pp. 41275–41296, March 2024. doi: 10.1007/s11042-024-20115-z.

[3] Y. Xiang, "Eye fatigue state recognition of Gabor wavelet optimization HMM algorithm," *International Journal of Hybrid Information Technology*, vol. 9, no. 9, pp. 189–198, 2016. doi: 10.14257/ijhit.2016.9.9.18.

[4] A. Kuwahara, R. Hirakawa, H. Kawano, K. Nakashi, and Y. Nakatoh, "Eye fatigue prediction system using blink detection based on eye image," in *Proceedings of the 2021 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, January 2021, pp. 1–4. doi: 10.1109/ICCE50685.2021.9427681.

[5] R. Ni, "Research on driver fatigue detection method based on convolutional neural networks using eye features," in *Proceedings of the IEEE Conference on Artificial Intelligence and Applications in High Performance Computing (AIAHPC)*, Wuhan, China, 2025, pp. 191–195.

[6] H. A. Madni, A. Raza, R. Sehar, N. Thalji, and L. Abualigah, "Novel transfer learning approach for driver drowsiness detection using eye movement behavior," *IEEE Access*, vol. 12, pp. 64765–64778, April 2024. doi: 10.1109/ACCESS.2024.3392640.

[7] L. Xie and S. Fan, "A learning-based multi-frame visual feature framework for real-time driver fatigue detection," in *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations)*, Albuquerque, NM, USA, April 2025, pp. 61–69. doi: 10.18653/v1/2025.naacl-demo.

[8] Y. Xiang, R. Hu, Y. Xu, C.-Y. Hsu, and C. Du, "Gaussian weighted eye state determination for driving fatigue detection," *Mathematics*, vol. 11, no. 9, Article 2101, April 2023. doi: 10.3390/math11092101.

[9] G. F. D. Almeida, *Real-Time Detection of Driver Fatigue Using Mobile Device Sensors and Artificial Intelligence On-Device*, M.S. Thesis, Department of Informatics Engineering, Faculty of Engineering, University of Porto, Porto, Portugal, July 2025.

[10] M. Vasiljevas, R. Damaševičius, and R. Maskeliūnas, "A human-adaptive model for user performance and fatigue evaluation during gaze-tracking tasks," *Electronics*, vol. 12, no. 5, Article 1130, February 2023. doi: 10.3390/electronics12051130.

[11] M. S. Chae *et al.*, "Eye-Blink and SSVEP-Based Selective Attention Interface for XR User Authentication: An Explainable Neural Decoding and Machine Learning Approach to Reducing Visual Fatigue," *IEEE Access*, vol. 13, pp. 176998–177018, 2025. doi: 10.1109/ACCESS.2025.3613355.

[12] S. Das, B. Bhaumik, B. Bagchi, S. Roy, and S. Mukherjee, "A comparative study on procedural and predictive systems on drowsiness detection," in *Proceedings of the 2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, Kolkata, India, April 2023, pp. 1–5. doi: 10.1109/IEMENTech60402.2023.10423474.

[13] S. Selvi, O. G. Elamathy, and S. Swetha, "Eye fatigue detection: A hybrid CNN–LSTM approach for digital screen users," in *Innovations in Computer Vision, Communication Systems, and Computational Intelligence*, S. L. Tripathi, O. P. Kumar, A. D. Stalin, and T. Ali, Eds. Boca Raton, FL, USA: CRC Press, 2025, pp. 75–81.

[14] H. Sidaq, L. Wang, S. Guizani, H. Haider, A. U. Rehman, and H. Hamam, "VMFD: Virtual Meetings Fatigue Detector Using Eye Polygon Area and Dlib Shape Indicator," *Computers, Materials & Continua*, vol. 86, no. 3, p. 25, 2026. doi: 10.32604/cmc.2025.071254.

[15] M. El Harti, S. J. Andaloussi, and O. Ouchetto, "A comprehensive review comparing artificial intelligence and clinical diagnostic approaches for dry eye disease," *Diagnostics*, vol. 15, no. 23, Article 3071, November 2025. doi: 10.3390/diagnostics15233071.

[16] G. Kiss and P. Viktor, "Artificial intelligence-based depression detection," *Sensors*, vol. 26, no. 2, Article 748, January 2026. doi: 10.3390/s26020748.

[17] C. F. Blanco-Díaz *et al.*, "Effects of the concentration level, eye fatigue and coffee consumption on the performance of a BCI system based on visual ERP-P300," *Journal of Neuroscience Methods*, vol. 382, Article 109722, October 2022. doi: 10.1016/j.jneumeth.2022.109722.

[18] I. Kayadibi, G. E. Güraksın, U. Ergün, and N. Öz. Süzme, "An eye state recognition system using transfer learning: AlexNet-based deep convolutional neural network," *International Journal of Computational Intelligence Systems*, vol. 15, Article 49, 2022. doi: 10.1007/s44196-022-00108-2.

[19] C. Wang, M. Lin, L. Shao, and J. Xiang, "Temporal facial features based fatigue detection system for low-cost devices," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 239, no. 14, pp. 7272–7286, February 2025. doi: 10.1177/09544070241306160.

[20] Q. Yang, Z. Li, C. Li, and C. Zhu, "Eye state recognition algorithm GAHMM of web-based learning fatigue," *Journal of Multimedia*, vol. 9, no. 3, pp. 397–405, 2014. doi: 10.4304/jmm.9.3.397-405.

[21] H. Rahman, M. U. Ahmed, S. Barua, P. Funk, and S. Begum, "Vision-based driver's cognitive load classification considering eye movement using machine learning and deep learning," *Sensors*, vol. 21, no. 23, Article 8019, December 2021. doi: 10.3390/s21238019.

[22] J. Lv and Y. Pei, "Driver fatigue prediction method based on BiLSTM and novel comprehensive fatigue index," *Human-Centric Intelligent Systems*, vol. 5, pp. 545–558, March 2025. doi: 10.1007/s44230-025-00113-6.

[23] Z. Lian, T. Xu, Z. Yuan, J. Li, N. V. Thakor, and H. Wang, "Driving fatigue detection based on hybrid electroencephalography and eye tracking," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 11, pp. 6568–6580, November 2024. doi: 10.1109/JBHI.2024.3446952.

[24] W. Sun, Y. Wang, B. Hu, and Q. Wang, "Exploration of eye fatigue detection features and algorithm based on eye-tracking signal," *Electronics*, vol. 13, no. 10, Article 1798, May 2024. doi: 10.3390/electronics13101798.

[25] A. Kuwahara, K. Nishikawa, R. Hirakawa, H. Kawano, and Y. Nakatoh, "Eye fatigue estimation using blink detection based on Eye Aspect Ratio Mapping (EARM)," *Cognitive Robotics*, vol. 2, pp. 50–59, 2022. doi: 10.1016/j.cogr.2022.01.003.